Response to Referee 2:

We would like to thank the referee for their insightful and very helpful comments that have improved the clarity of this manuscripts. We would like to address their latest points below.

1) The so-called synthetic observations outline in lines 245-250 of the revised manuscript ("The synthetic observations were generated using 5000 rays, etc.") need in my opinion some clarifications in the text:

a) Does that mean that 5000 heights of synthetic aircrafts were generated?

Yes. This has been clarified in the main text as "The synthetic observations (heights of synthetic aircraft) were generated using 5000 rays…".

b) Starting from the prior given by Eq. (23), were 5000 synthetic refractivity profiles calculated? In that case, how were the profiles of fig. 8a obtained? Perhaps averaging the 5000 profiles? Was the 0.00° RMS profile obtained from the 5000 interferometrically-measured observed AoAs. Were the 0.01° and the 0.05° RMS profiles obtained by perturbing the 5000 interferometrically-measured observed AoAs according to the corresponding normal distributions?

The following has been added on line 280 for clarification:

"The 0.00 deg. RMSE profile was obtained from the 5000 interferometrically-measured, unperturbed observed AoAs. The 0.01 deg. and the 0.05 deg. RMSE profiles were obtained by perturbing the 5000 interferometrically-measured observed AoAs according to the corresponding normal distributions."

c) I don't understand the final sentence in the caption of fig. 8: "The error bars show the range in retrieved refractivity profiles using twenty different observed AoA noise distributions." What is understood by "noise distribution" in this context? Do the authors mean "noise realizations" instead of "noise distributions"? Wouldn't it be more rigorous to compute (and represent) the standard deviation of the noisy profiles?

Yes, this should be noise realisations. The fractional error standard deviation is included in Fig 8b for the twenty noisy retrievals. The following has been added to line 284, describing the bias increasing as the AoA noise increases:

"The 0.05 deg. RMSE profile shows a lower refractivity error standard deviation at higher altitudes than the 0.01 deg. RMSE profile. The increased standard deviation in the AoA measurement error results in a greater number of AoAs being perturbed below 0 deg. for the 0.05 deg. RMSE case. Since the retrievals can currently only be performed using AoAs ≥ 0 deg., these are then rejected, resulting in a bias towards perturbing the AoAs towards higher elevation angles. This results in the gradient in the retrieved refractivity profile becoming steeper which, as a

consequence of the clamping of minimum refractivity, results in the 0.05 deg. RMSE profiles clustering together with a lower variance. However, the larger RMSE in the observed AoA results in a larger bias in the retrieved refractivity profile."

2) I still have concerns on the message figures 9a and 9b are meant to convey. On the one hand, the synthetic retrieved refractivity profiles of fig. 8 fit quite well the radiosonde-derived profile and one can think that this is because, despite they have been calculated on the synthetic aircraft positions, these positions were generated in turn using the radiosonde refractivity profile. On the other hand, no retrieved synthetic profiles are shown in fig. 9, just the radiosonde-derived profiles ("The figures for the synthetic refractivity retrievals using the 15 December and 18 July 2022 refractivity profiles are not shown") and one has to make do with the assertions that, even injecting noise, there is improvement with respect the initial guess profile, but that, under noise with 0.01° (or greater) standard deviation, the positive refractivity gradient in fig. 9a cannot be resolved. To this point, I don't see the information added by the figures beyond the fact that negative refractivity gradients can happen. In addition, if I'm not wrong, the observed AoAs from 22 September 2023 were also used (at least there is no mention to other AoA observations) to generate synthetic retrievals in these two different days, several months earlier. To what extent is this meaningful? Is it sensible to use observed AoAs obtained in some conditions of refractivity profile to generate synthetic aircraft positions in (very) different conditions to test the retrieval algorithm in the latter conditions and trying to obtain valid information? Can it be excluded that the failure to retrieve profiles with negative refractivity gradients in the presence of injected noise is caused by this mismatch between the observations obtained under a given profiles and the synthetic aircraft positions being calculated under a potentially very different profile? All this in turn prompts a general question about the purpose of the synthetic retrievals: are they just intended to test the algorithm and its robustness against noise? Does this have an impact on the "real" retrievals of fig. 11?

We thank for the referee for their helpful suggestions. We have removed Fig. 9, showing the radiosonde profiles (but no retrievals) and instead opted to just describe the results in the main text and in Table 1. The synthetic retrievals for the 18 July and 15 Dec were obtained using an identical process to that used to generate the 22 Sep 2023 data. The respective profiles were used to obtain the "true" synthetic aircraft heights, then noise was added to the AoAs and the original profiles were retrieved using these noisy AoAs and synthetic aircraft heights. A description has been added in the manuscript in the last paragraph of section 3.1:

"The synthetic refractivity retrievals were then repeated for two other radiosonde profiles - the 12:00 UTC Watnall soundings for the 18 July 2022 and 15 December 2022 respectively. The synthetic retrievals were performed using synthetic observations of aircraft heights generated from the two respective refractivity profiles in a process identical to that using the 22 September data."

The following was also added, to summarise the reasoning behind the synthetic retrievals:

“The synthetic refractivity retrievals demonstrate the ability to obtain useful atmospheric data using measurements of refracted ADS-B transmissions in the presence of random noise.”

3. Connected to the last question, I think that figure 11, or its context, could be more informative. For example, how many observations were used to generate the profile shown for each time slot? If, as I suppose, several observations were used, how is the shown profile obtained (see question 1b). Could the variability of the profiles generated from observations within a time slot be used to assess the uncertainty of the final profile? Likewise, how many observations have been used in the graphs of figure 12 for each of the time slots. Are they the same used to generate the profiles of fig. 11?

The following statement was added to the first paragraph of section 3.2:

“The four observational periods contained 9700 (08:36-08:51 UTC), 8315 (08:51-09:06 UTC), 8185 (09:06-09:21 UTC) and 7589 (09:21-09:36 UTC) received ADS-B transmissions respectively within the sector.”

The number of observations in each observational period was also included in the caption for Fig. 10.

In terms of estimating the refractivity uncertainty in the retrievals, this is currently challenging to estimate. The retrievals require a suitable distribution in aircraft heights (since it is a tomographic problem), and neglecting certain aircraft heights has a significant impact on the retrieval quality. Future experiments plan to assimilate the data into numerical weather prediction models (rather than direct retrievals of vertical profiles) using either a variational approach or through the use of ensemble Kalman filters. The observations for the four periods were the same for Figs 9, 10 and 11 (stated in main text).

4. The main source of uncertainty in the observations is attributed to multipath (lines 259-261: "Since multipath contamination of the observed AoA measurement is thought to constitute the principal source of uncertainty in the retrieved refractivity profile, three observed AoA measurement noise test cases were investigated"). Can these multipath effects be mimicked by normally-distributed noise? In case it is justified they can, what's the amount of noise to be expected for the real retrievals?

This is currently outside the scope of the paper, as the multipath environment is extremely complex. Future experiments will likely require wave optics modelling to determine the uncertainty arising due to multipath. Some AoAs are also biased, which is difficult to account for in the retrievals currently. A statement has been added to the end of section 4.3:

“Future experiments using the prototype ADS-B interferometer will aim to quantify the nature of the observational noise arising due to multipath contamination.”

Minor remarks:

1. Is it necessary the redundancy between what is said in lines 227-228 of the revised manuscript (“The synthetic and real retrievals were obtained using the central 10◦ azimuthal sector of received transmissions”) and the sentence in lines 234 and 235 (“ The observations used in the analysis were restricted to the central 10° azimuthal sector of the entire observation dataset”)?

This has been corrected, including just the first statement.

2. I suggest to modify the writing at the beginning of line 245 from “the reported distances of each of the ADS-B transmission” to “the reported distances s_T (see fig. 3) of each of the ADS-B transmission”.

This has been added.

3. I would suggest to modify the sentence in the last two lines of page 13 of the revised manuscript, which currently reads “A prior (first-guess) refractivity profile was used to initialise the penalty function, given by”, to “A prior (first-guess) refractivity profile, given by

$$N(h)=N_0 \exp(-h/H), \quad (23)$$

was used to initialise the penalty function”. This is to make clear that Eq. (23) is the first-guess refractivity profile, not the penalty function.

This has been corrected.

4. Vertical axes of the panels in figs. 11 and 12: the labels should probably be "LoS AoA (deg.)" or "AoA_Los" in view of the new terminology defined in lines 312-316.

This has been corrected.

5. Line 433: "The convergence of the penalty function was impacted negatively and the RMS in the retrieved refractivity profile". There is a noun (probably "error") missing after RMS, which acts as an adjective.

This has been corrected, and throughout has been corrected to RMSE (rather than RMS).